AMENDMENTS TO THE SPECIFICATION

Please replace the paragraph beginning at page 3, line 5, and insert the following rewritten paragraph:

-- In one of the conventional chorus detection methods, one chorus section of a specified length is incompletely extracted as a representative part of audio signals of a piece of music. Logan, B. and Chu, S., Music Summarization Using Key Phrases, Proc. Of ICASSP 2000, II-749-752 (2000),Logan, et. al [Prior Art 1] proposed a method of labeling a short extracted frame (1 second) based on acoustic features thereof, wherein a frame having the most frequent label is considered as a chorus. The labeling utilized clustering based on similarity in acoustic features among respective sections, or hidden Markov model. Bartsch, M. A. and Wakefield, G. H., To Catch A Chorus: Using Chroma-based Representations for Audio Thumbnailing, Proc. of WASPAA 2001, 15-18 (2001),Bartsch, et. al [Prior Art 2] proposed a method of dividing a piece of music into short frames for every beat based on the result of beat tracking, and extracting a part, as a chorus, which has the highest similarity of acoustic features thereof across sections of a certain specified length. Foote, J., Automatic Audio Segmentation Using a Measure of Audio Novelty, Proc. of ICME 2000, I-452-455 (2000),Foote [Prior Art 3] pointed out a possibility that a chorus can be extracted, as an application of detecting

a boundary based on similarity in the acoustic features among very

short fragments (frames). --

Please replace the paragraph beginning at page 3, line 23, and insert the

following rewritten paragraph:

-- Although there are the prior art intended for expression

equivalent to musical notes such as a standard MIDI file, etc. [Prior

Arts 4 and 5](e.g., Meek, C. and Birmingham, W. P., Thematic

Extractor, Proc. of ISMIR 2001, 119-128 (2001); and Jun Muramatsu,

Extraction of Features in Popular Songs Based on Musical Notation

Information of "Chorus"--Case of Tetsuya Komuro, The special Interest

Group Note of IPSJ, Music Information Science, 2000-MUS-35-1, 1-6

(2000)), this technology could not be directly applied to mixed sounds

wherein it was difficult to separate sound sources. The conventional

chorus section detecting method could simply extract and present

sections of a certain specified length at any given time, and could not

estimate where the chorus sections begin and end. Furthermore, no

prior art have taken modulation into consideration. --

Please delete the paragraphs from page 4, line 6, to page 5, line 18.

Please replace the paragraph beginning at page 54, line 7, and insert the

following rewritten paragraph:

-- If no selection buttons have been clicked in step ST12, it goes to step ST14. In step ST14, it is judged whether or not an operation has been performed giving instructions to move the mark 'a' by clicking (touching) the mark 'a' of the playback slider. If selection buttons have been clicked in step ST12, it goes to step ST13. In step ST14, the playback position is changed to the beginning of the clicked section. If the operation has been performed, it goes to the step ST15, where the playback position is set to the position where the mark 'a' of the slider has moved, and then it returns to step ST2, after setting the playback condition to playback in step ST7. --

Please replace the paragraph beginning at page 54, line 26, and insert the following rewritten paragraph:

-- In Fig. 6 and Fig. 7, the "playback condition" includes stopped, paused, and playing conditions; the "playback position" refers to the elapsed time from the beginning of the file of the piece of music; and the "playback speed" includes normal playback speed, fast forward playback speed, and fast reverse playback speed. In Fig. 7, in step ST21, the playback condition has been set to STOP. Then, in step ST22, the playback position has been set to the very beginning. And then, the step proceeds to step ST23, where the playback condition has been not PLAY, it returns to step ST23, where the playback condition has been PLAY, the step proceeds to step ST24. In step

ST24, it is judged whether or not the playback position has been at the

end of the file. If the playback position has been at the end of the file, it

returns to step ST21. If not, it goes to step ST25. In step ST25, it is

judged whether or not the playback speed is fast-forwarding or

rewinding. If the playback speed has been fast-forwarding or

rewinding, it goes to step ST28. If not, it goes to step ST26. In step

ST26, a very short section from the playback position is played back at

normal speed. Then, the playback position is updated in step ST27

and then it returns to step ST23. And, in step ST28, the very short

section from the playback position is played back at fast-forwarding or

rewinding speed. Then, the playback position is updated in step ST29

and then it returns to step ST23. --

Please replace the paragraph beginning at page 64, line 10, and insert the

following rewritten paragraph:

-- (3) The similarities between the acoustic features of the

extracted 12-dimensional chroma vectors and that of all previous

frames are calculated (solution to Problem 1) (Step S3-1). Then, pairs

of repeated sections are listed while automatically changing the

repetition judgment criterion for every piece of music, by adopting the

Automatic Threshold Selection Method (Noriyuki Otsu, Automatic

Threshold Selection Method Based on Discrimination and Least

Square Criterion, Journal of Institute of Electronics, Information and

Communication Engineers (D), J63-D, 4, 349-356 (1980))[Prior Art 6]

based on the judgment criterion (solution to Problem 2) (Step S3-2).

Then, create a group of repeated groups by integrating those pairs

over the entire piece of music, and also determine respective end

positions appropriately (solution to Problem 3) (Step S3-3). --

Please replace the paragraph beginning at page 66, line 15, and insert the

following rewritten paragraph:

-- Now a 12-dimensional chroma vector is described with

reference to FIGS. 18 and 19. A chroma vector is an acoustic feature

representative of power distribution, with a chroma (pitch class,

chroma) disclosed in Shepard, R. N., Circularity in Judgments of

Relative Pitch, J. Acoust. Soc. Am., 36, 12, 2346-2353 (1964), the prior

art 7 as a frequency axis. A chroma vector here is close to what results

from scattering the chroma axis of the chroma spectrum shown in the

prior art 8 into 12 pitch classes. As shown in FIG. 18, perception of

musical pitches (musical height and pitch height) according to the prior

art 7 Shepard has an ascending helical structure. And the perception of

musical pitches can be expressed in two dimensions: a chroma on the

circumference when the helix is viewed from above, and longitudinal

height (an octave position, height) when it is viewed from the side. For

a chroma vector, considering that the frequency axis of the power

spectrum runs along the helical structure, and by crushing the helix into

a circle in the direction of height axis, the frequency spectrum is

expressed only by the chroma axis on the circumference (one

perimeter constitutes one octave). In other words, powers of positions

under a same pitch class over different octaves are added to make the

power of the position of the pitch class on the chroma axis. --

Please replace the paragraph on page 76, line 1 to page 76, line 10, with the

following rewritten paragraph:

-- An adequately high peak sin $R_{all}(t,1)$ as shown in the diagram

on the right side of FIG. 28 are detected as a line segment candidate

peaks. First, peaks of $R_{all}(t,l)$ relative to the lag axis are determined by

the peak detection using smoothing differentiation through matching of

quadratic polynomials (Savitzky, A. and Golay M. J., Smoothing and

Differentiation of Data by Simplified Least Squares Procedures,

Analytical Chemistry, 36, 8, 1627-1639 (1964)) [prior art 9].

Specifically, a location where the smoothing differentiation of $R_{all}(t,l)$

determined by the following equation (8) changes from positive to

negative shall be considered a peak ($Ks_{ize}=0.32_{sec}$). --

Please replace the last paragraph beginning on page 76 and spanning to

page 77, line 10, with the following rewritten paragraph:

-- Next, from a collection of thus obtained peaks, only peaks that are higher than a certain threshold are selected as line segment candidate peaks. As discussed in the Problem 2 mentioned earlier, the threshold should be automatically changed based on a piece of music because an appropriate value differs for every piece of music. Thus, when the peak heights of $R_{all}(t,l)$ are dichotomized into 2 classes by a threshold, the automatic threshold selection method [prior art 6](Noriyuki Otsu, Automatic Threshold Selection Method Based on Discrimination and Least Square Criterion, Journal of Institute of Electronics, Information and Communication Engineers (D), J63-D, 4, 349-356 (1980)) is used based on the discrimination criterion providing a maximum degree of class separation. As shown in FIG. 29, the automatic threshold selection method has adopted the idea of dichotomizing the peak heights into two classes by a threshold. Here, as the degree of class separation, a threshold that maximizes inter-class distribution is determined; --

Please replace the paragraph beginning at page 80, line 26, and insert the following rewritten paragraph:

-- The integrated repeated sections determined by the integrated repeated section determination means 113 are stored in the integrated repeated section storing means 115 as integrated repeated section

rows. Fig. 35 shows an example in which the integrated repeated section rows are displayed in a display means 118 shown in Fig. 6. --

Please replace the paragraph beginning at page 94, line 22, and insert the following rewritten paragraph:

-- As an evaluation experiment, detection capability of this apparatus was examined for 100 pieces of music (RWC-MDB-P-2001, No.1 to No.100) of the popular-music database "RWC Music database: Popular Music" [Prior Art 10](Masataka Goto, Hiroki Hashiguchi, Takuichi Nishimura, and Ryuichi Oka, RWC Music Database for studies; Popular Music Database and Copyright-Expired Music Database, The special Interest Group Note of IPSJ, Music Information Science, 2001-MUS-42-6, 35-42 (2001)). When a whole piece of music has been entered, what are detected as chorus sections are evaluated. In addition, to provide a reference for judging whether detection results are right or wrong, correct chorus sections have to be labeled manually. To enable this task, a music structure labeling editor was developed, which can divide up the piece of music, and label each section as a chorus, verse A, verse B, interlude, etc. In the labeling, the relative width of key shift (by how many semitones the key is shifted up with respect to the key at the start of the music) is also taken into consideration to determine correct chorus sections. --

Please replace the paragraph from page 95, line 11 to page 95, line 16, and insert the following rewritten paragraph:

-- Based on thus created reference of the correct chorus sections, the degree of matching between the detected and the correct chorus sections was evaluated in terms of the recall rate, the precision rate, and the F-measure, which is the harmonic mean thereof ~~(Prior Art 11)~~) van Rijsbergen, C. J., Information Retrieval, Butterworths, second edition (1979)). The definition is shown below: --

Please replace the paragraph beginning at page 97, line 24, and insert the following rewritten paragraph:

-- Furthermore, the present invention is also related to music summarization ~~[prior art 12]~~(Keiji Hirata, Shu Matsuda, Papipoon: GTTM-based Music Summarization System, The special Interest Group Note of IPSJ, Music Information Science, 2002-MUS-46-5, 29-36 (2002)), and the apparatus of the present invention can be regarded as a method of summarizing a piece of music which presents chorus sections as the result of the summarization. In addition, when a summary of a section longer than a chorus section is needed, the use of repetition structures, which have been acquired as the intermediate results, enables to present a summary reducing the redundancy of an entire piece of music. For instance, when a repetition of (verse A $\rightarrow$

verse B → chorus) is captured as the intermediate result, it can be

presented. --